1245689

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

November 04, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/492,965*
FILING DATE: *August 07, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/22314*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

PROVISIONAL APPLICATION COVER SHEET

U.S. Patent and Trademark Office
[illegible] South Clark Place
[illegible]omer Window, Mail Stop Provisional Patent Application
[illegible]tal Plaza Two, Lobby, Room 1B03
[illegible]ington, Virginia 22202

This is a request for filing a PROVISIONAL APPLICATION under 37 C.F.R. § 1.53(c).

Docket Number	ATI-0029PR	Type a plus sign (+) inside this box -	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Van Bokkelen	Gil	Bradford	Cleveland Heights, OH
Basu	Joydeep		Shaker Heights, OH

TITLE OF THE INVENTION (280 characters max)

ENHANCED SYNTHETIC MICROCHROMOSOME FORMATION FROM ALPHA SATELLITE WITH ARTIFICIALLY INCREASED DENSITY OF CENP-B BOXES

CORRESPONDENCE ADDRESS

Please direct all correspondence to Customer Number **34610**

STATE	Virginia	ZIP CODE	20153-1200	COUNTRY	USA

ENCLOSED APPLICATION PARTS *(check all that apply)*

[X]	Specification	*Number of pages* [28]	[X]	Applicant claims Small Entity Status
[X]	Drawings	*Number of sheets* [5]	[ ]	Other (specify):

METHOD OF PAYMENT *(check one)*

[X]	A check or money order is enclosed to cover the Provisional filing fees	FILING FEE	
[ ]	The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 16-0607	[ ] $160.00 [X] $80.00	$80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the name of the U.S. Government agency and the Government contract number are: _____

[ ] Additional inventors are being named on separately numbered sheets attached hereto.

Respectfully submitted,

FLESHNER & KIM, LLP

Steven J. Helmer

Donald R. McPhail
Registration No. 35,811
Steven J. Helmer
Registration No. 40,475

P.O. Box 221200
Chantilly, Virginia 20153-1200
703 502-9440
Date: August 7, 2003
DRM\SJH:dcp

Please direct all correspondence to Customer Number 34610

Enhanced synthetic microchromosome formation from alpha satellite with artificially increased density of CENP-B boxes

Joydeep Basu*, [redacted], Gwen Hau, Jessica Dolhonde, [redacted] and Gil Van Bokkelen

Athersys, Inc., 3201 Carnegie Avenue, Cleveland, OH 44115, USA

Tel (216) 431 9900 Fax: (216) 361 9596

e-mail: jbasu@athersys.com

* Author for correspondence

Key words: alpha satellite, centromere, CENP-B, synthetic microchromosome, human artificial chromosome

ABSTRACT

The presence of binding sites for the centromere protein CENP-B (the 'CENP-B box') has been correlated with the ability of alpha satellite DNA to form centromeres *de novo* in synthetic microchromosome (SMC) assays. However, the effect of the density of CENP-B boxes on the frequency of SMC formation has not previously been explored. Here, we report a systematic analysis of the role of the CENP-B box in human alpha satellite DNA, using the formation of SMCs as an assay for the establishment of centromere function. We have created synthetic alpha satellite arrays, based on the 16-monomer repeat length typical of natural chromosome 17-derived D17Z1 arrays. In these synthetic arrays, the consensus CENP-B box elements are either completely absent (0/16 monomers) or are increased in density (16/16 monomers) compared to D17Z1 alpha satellite (5/16 monomers). We show that, not only is the presence of CENP-B box elements a requirement for efficient *de novo* centromere formation, but that increasing the density of CENP-B box elements results in an enhancement of the efficiency of *de novo* centromere formation. These findings have implications for the design of strategies to construct novel SMC vectors for functional genomics and potential therapeutic applications.

INTRODUCTION

Alpha satellite DNA is the major species of repetitive element found at the centromeres of all normal primate chromosomes. It is organized in a hierarchical structure based on a ~171 bp monomeric unit that is multimerized in a tandem manner into a higher-order repeat, which is further multimerized over hundreds of kilobases at the centromeres of all normal human chromosomes (reviewed in 1, 2, 3 ,4). Centromeric alpha satellite acts to organize the recruitment of key

centromeric proteins (CENPs) to form a trilaminar protein/DNA complex, the kinetochore, which mediates the interactions between the chromosome and the spindle apparatus that are responsible for coordinated chromosome movements during cell division (5). While functional kinetochores have been observed at chromosomal locations not containing any alpha satellite (so called "neo-centromeres"; reviewed in (6)) only cloned alpha satellite DNA has thus far been shown to form centromeres *de novo* when introduced into the cell nucleus by transfection or microinjection in synthetic microchromosome (SMC) or artificial chromosome assays (7, 8, 9).

The ability to create human SMCs was pioneered through the development of techniques to synthesize megabase-sized alpha satellite arrays *in vitro* (10), starting with a single cloned copy of a higher-order repeat (11). These SMC vectors may have potential applications in human gene transfer (7,12); for example, SMCs containing the *HPRT* genomic locus have been shown to complement HPRT-deficient cell lines (Rudd et al., in press; 13,14), and we have observed sustained expression of the β-globin gene from SMCs carrying the entire 150 kb β-globin genomic region (Basu et al., in preparation). In addition, SMC and artificial chromosome vectors provide a methodological platform for the identification and functional analysis of elements in alpha satellite that are critical for centromere function (Rudd et al., in press; 15, 16, 17, 10).

Variations in the efficiency of *de novo* centromere formation between alpha satellite templates derived from different human chromosomes (18, 8, 16) have demonstrated a causal link between the presence of sequence elements called CENP-B boxes and *de novo* centromere seeding efficiency (15, 19). The CENP-B box is the biochemically-defined motif "PyTTCGTTGGAAPuCGGGA" minimally responsible for mediating binding of the constitutive centromere protein CENP-B to human alpha satellite DNA (20, 21).

While there is clear evidence implicating the presence of CENP-B boxes in *de novo* centromere formation (15), it is not clear to what extent the density of CENP-B boxes might influence the efficiency of SMC formation. Thus, in order to address the functional significance of the CENP-B box in human alpha satellite and in SMC formation, we have developed methodologies to directly vary the density and distribution of CENP-B boxes in the D17Z1, chromosome 17-derived HOR, which in its natural configuration contains a CENP-B box in 5 of its 16 constituent monomers. We have constructed entirely synthetic D17Z1 HOR derivatives, in which each of the 16 tandem monomeric repeats contains either a consensus CENP-B box or a related sequence element derived from Y chromosome alpha satellite, which does not bind CENP-B (22, 23). Here, we report that the efficiency of formation of SMCs is directly proportional to the density of CENP-B boxes in the SMC vector, thus demonstrating a requirement for CENP-B boxes in centromeric chromatin assembly. As the methods we present here are generally applicable, these data have implications for the design and further development of SMCs for potential applications in human gene therapy.

MATERIALS AND METHODS

Synthesis of modified 2.7 kb chromosome 17-derived higher-order repeats

The sequence of the 2.7 kb D17Z1 higher-order repeat (11) was modified such that each of the 16 monomer units contained the consensus CENP-B box element 5': TTT CGT TGG AAA CGG GA: 3' (22) or the related Y alpha satellite-derived element AGA TGG TGG AAA AGG AA, which lacks CENP-B-binding activity ('CENP-B box null'). Each of the 16 modified monomer units was then synthesized by ligation of two to three pairs of overlapping oligonucleotides (Operon Technologies, CA). Adjacent pairs of mutated monomer

units were then ligated together to form dimers. In addition, the EcoRI sites of monomers 1 and 16 were altered to create a BamHI site at the 5' end of monomer 1 and a BglII site at the 3' end of monomer 16. Each gel-purified dimer was then PCR amplified with a BsaI or SapI restriction site, such that upon digestion each dimer would produce a defined overhang exactly complementary to an overhang in the adjacent dimer. The resultant tetramers (containing no extraneous sequence) were then T/A subcloned into pGem-Teasy (Promega) and sequence verified. Adjacent tetrameric subunits were then ligated together using SapI (or NotI and SapI for monomers 1 and 16) to generate the appropriate overhang. The resultant octamers were further gel purified and ligated together to produce the completed synthetic 16-mer, representing a single D17Z1 higher-order repeat unit, with NotI overhangs. This higher-order repeat was then subcloned as a NotI fragment into the BAC cloning vector pBeloBAC11 (24). The overall strategy is outlined in Figure 1A.

Directional multimerization of the synthetic higher-order repeats

The 2.7 kb CENP-B box enriched or CENP-B box null D17Z1 higher-order repeat was multimerized directionally as follows. The cloned synthetic higher-order repeat (in pBeloBAC11) was digested with BamHI and SpeI, and this band (fragment 'A') was gel purified by standard procedures (Qiagen). A second fragment ('B') was generated by digesting the same cloned repeat with BglII and SpeI. The appropriate fragment 'B' was subsequently gel purified and ligated to the BamHI/SpeI digested fragment 'A'. This ligation reaction was transformed into *E.coli* (GibcoBRL), and recombinant clones identified by NotI digestion of the resultant clones and pulsed field gel electrophoresis (Fig. 1B). This process was repeated iteratively to create clones containing 4, 8, 16 and 32 copies of the CENP-B box enriched/CENP-B box null chromosome 17 based higher-order repeat in

pBeloBAC (Fig. 1C). Finally, for use as a selectable marker in mammalian cells, a cDNA cassette conferring resistance to puromycin was introduced into 17α32(CENP-B box enriched/null) unit/pBeloBAC by transposition of the puroR cassette into the pBeloBAC vector backbone (Epicentre).

An ~86 kb synthetically assembled alpha satellite array, derived from directional multimerization of the *naturally* occurring 2.7 kb D17Z1 repeat unit (p17H8, see 8, 10, 11), was subcloned as a BamHI/BglII fragment into the BamHI site of pBeloBAC11. This construct, 17α32(natural)/pBeloBAC, was further modified by transposition with a puromycin resistance selectable marker (Epicentre). The structural integrity of all modified higher-order repeats and of the original higher-order repeat array was confirmed by sequencing, restriction digestion and FISH hybridizations using the array as probe.

Mobility shift analysis

The effect of mutations described above on CENP-B binding to the synthetic HOR was evaluated by a gel mobility shift assay. Cloned tetramer units assembled from CENP-B box-enriched and CENP-B box-null monomers were digested with NotI and inserts were gel purified. Subsequent to incubation with purified recombinant CENP-B protein (Diarect, Germany) for 25 minutes at room temperature in CENP-B binding buffer (20), protein/DNA complexes were electrophoresed through a 2% agarose gel in 0.5xTBE buffer. Following electrophoresis, SybrGold (Molecular Probes) stain was used to visualize DNA bands.

Cell transfection

Human HT1080 cells (gift of Dr. Brenda Grimes, Case Western Reserve University) were transfected using the Fugene 6 (Roche) reagent according to the

manufacturer's instructions, and stable clones identified on the basis of resistance to puromycin (Kayla) at 3 μg/ml. Clones appeared after 7-10 days and were subsequently expanded to generate clonal lines for further analysis.

Cytogenetic analysis and validation of SMCs

Clonal populations of cells containing potential SMCs were analyzed, generally as described (8, 16, 10). Briefly, cells were arrested at metaphase using colchicine (Gibco) at 40 ug/ml for 45 minutes at 37 degrees Celsius, then treated with hypotonic solution (0.075 M KCl, 12 minutes, 37 degrees Celsius) and applied to slides using the Shandon Cytospin 3. Slides were subsequently fixed in 2% formaldehyde solution and immunoreacted with rabbit anti-CENP-C antibody (10) at a concentration of 1/2000 in PBS and detected with goat anti-rabbit IgG (H + L) (Molecular Probes). DNA probes were labeled by nick translation using the Vysis system according to the manufacturer's instructions. Immunoreacted slides were fixed (3:1, methanol:acetic acid), subjected to denaturation (70% formamide, 72 degrees Celsius, 8 minutes), and hybridized to denatured probes as described (8).

Putative artificial chromosomes were scored if they showed a positive hybridization signal with a FISH probe derived from the synthetic array as well as positive CENP-C immunoreactivity. Mitotic stability was evaluated by growth in the absence of drug selection for up to six weeks.

RESULTS

Previous studies have established that vectors containing multiple copies of certain alpha satellite higher-order repeat units can seed formation of *de novo* centromeres in human HT1080 cells (8, 10, 15-18; Rudd et al., in press). However, the overall frequency of generation of SMCs has been reported to be

quite variable and often quite low (Rudd et al., in press; 25, 15, 8, 18), depending at least in part on the chromosomal origin of the alpha satellite array and on the presence or absence of CENP-B boxes. Therefore, we have undertaken to develop a general approach to maximize the efficiency of SMC formation and to evaluate the sequence-dependency of *de novo* centromere seeding.

Construction of Engineered, D17Z1-based higher-order repeats

The SMC system provides a platform to systematically evaluate the functional significance of sequence elements within human alpha satellite DNA. We developed methodologies to construct modified synthetic D17Z1 units that are either enriched or depleted in the density of CENP-B box DNA binding elements. The higher-order repeat unit of D17Z1 alpha satellite consists of 16 monomer units (11). In order to generate engineered higher-order repeats, each of the 16 monomer units was synthesized by the serial stepwise assembly of oligonucleotide pairs, each between 60 and 100 bp in length, as shown in Figure 1A. Adjacent monomer units could then be gel-purified and ligated to form dimers. Each dimer was PCR-amplified to introduce a restriction site such as SapI (which cuts outside its recognition sequence and can generate custom-made overhangs that can be ligated seamlessly), thereby generating tetramers without the addition of any extraneous sequence. This process of PCR and ligation assembly was serially repeated until the complete 16-mer repeat unit was constructed. The resulting synthetic higher-order repeat was then subcloned and directionally concatamerized to 32 copies (Figure 1B, C), using methods previously developed in our laboratory (10).

CENP-B boxes are required for efficient centromere formation *de novo*

We used the approach described above to create a modified variant of D17Z1 alpha satellite in which all the consensus CENP-B boxes or elements resembling the consensus in each of the 16 monomer units were replaced with a sequence derived from Y chromosome alpha satellite. This approach allowed us to knockout any interaction between CENP-B and its biochemically defined consensus element, as well as any interactions between CENP-B and elements resembling the consensus that might potentially occur *in vivo*. Confirmation of abolishment of CENP-B binding to the synthetic CENP-B null array was shown by loss of mobility shift in a gel shift assay (Figure 2).

Constructs based on the naturally occurring, unmodified D17Z1 have been used previously to generate mitotically stable SMCs in greater than 10% of drug-resistant clones after transfection into human HT1080 cells (Rudd et al., in press; 8, 10, 18). Here, SMCs were identified in 4 of 38 colonies (Table 1), consistent with earlier data. However, when using the CENP-B null construct in which all CENP-B boxes had been modified, only a single clone was identified to have a putative SMC out of 40 clones screened, representing a maximum *de novo* centromere formation frequency of 2.5 % (Table 1). The fact that the observed rate of *de novo* SMC formation is low but is not zero is consistent with other reports that some alpha satellite arrays that do not contain CENP-B boxes can in fact mediate apparent SMC formation at very low frequencies (25, 18), although the possibility that these represent SMCs that have acquired endogenous centromere sequences has not been rigorously excluded. Indeed, previous data have demonstrated that the likelihood of such an acquisition event is increased when the *de novo* centromere competency of the transfected DNA is lowest, as in the case of CENP-B null constructs (8, Rudd et al., in press). Our data are in agreement with those recently reported by Masumoto and colleagues, who used a similar approach to abolish CENP-B boxes in a higher-order repeat derived from chromosome 21

(15). Combined, the two studies provide strong evidence that CENP-B boxes are required generally for efficient formation of *de novo* centromeres in SMC systems.

Creation of more efficient centromere constructs by increasing the density of CENP-B boxes

Several studies have now suggested a relationship between the presence of CENP-B boxes in cloned alpha satellite and the ability to form *de novo* centromeres from BAC or YAC vectors containing the cloned arrays (8, 10, 15-19). As an extension of the data presented above and by Ohzeki et al. (15), we reasoned that if the density of CENP-B boxes was indeed critical for *de novo* centromere formation, it might be possible to create synthetic alpha satellite arrays with a CENP-B box density even higher than their naturally occurring counterparts. These novel synthetic arrays might form a more efficient template for centromere formation *de novo* than natural arrays.

To evaluate this hypothesis, we used the strategy described above to construct a synthetic D17Z1-derived alpha satellite array supersaturated with CENP-B boxes, such that each of the 16 monomers in the HOR contained a consensus CENP-B box. Notably, upon introduction into HT1080 cells by transfection, these supersaturated synthetic arrays formed SMCs *de novo* more than twice as efficiently as arrays containing the natural density of CENP-B boxes (Table 1). The frequency of SMCs within any one clone was observed to vary from 10% to 100%, similar to the ranges observed in cell lines derived from transfection with the control natural arrays (8, 17). No integration events were observed cytogenetically, although Southern blot data (not shown) demonstrated the presence of BAC-specific DNA.

Consistent with other studies, cytogenetic estimates suggested that the SMCs (from all versions of the array) are several megabases in size. In all cases, SMCs

were shown to be mitotically stable in the absence of selection for six weeks and to bind the centromere-specific protein CENP-C. Thus, the effect of changing the density of CENP-B boxes appears to be limited to the efficiency of formation of SMCs and not to their subsequent behavior in mitotic segregation.

DISCUSSION

Since the original report of *de novo* centromere and SMC formation (10), a number of groups have described related approaches to further develop and optimize artificial chromosome systems (reviewed by 7, 26, 12). The creation of SMCs has now been established as a tractable approach to systematically identify and dissect elements that are critical for chromosome function (15, 16, Rudd et al., in press). In this report, we describe the further refinement of the SMC system as a methodological platform to undertake a functional analysis of the role of the density of CENP-B box elements in human alpha satellite DNA.

CENP-B is a constitutively present DNA-binding protein found in the underlying centric heterochromatin of all human chromosomes except the Y chromosome. The corresponding DNA sequence element that defines the cognate binding site, the CENP-B box, has been identified as PyTTCGTTGGAAPuCGGGA (20, 22) and is found distributed within some, but not all, of the monomer units of alpha satellite DNA from most human centromeres (25, 16, 27). However, the role of CENP-B if any, in specifying centromeric identity globally remains unsettled (28). Y chromosome centromeres do not associate with CENP-B (23), and African Green Monkey centromeres lack CENP-B boxes even though the CENP-B protein itself is present (29). Furthermore, Cenp-B knockout mice show only modest phenotypic effects and appear to have

fully functional centromeres as evidenced by the lack of chromosome missegregation phenotypes (30, 31, 32).

Notwithstanding this mechanistic uncertainty, studies of *de novo* centromere formation with cloned alpha satellite arrays support a direct correlation between the presence of CENP-B boxes and the competence of a construct for *de novo* centromere formation. For example, comparison of cloned alpha satellite arrays from chromosomes Y, X, 17 and 21 show that 17- and 21-derived arrays form *de novo* centromeres much more efficiently than X- and Y-derived arrays (Rudd et al., in press; 8, 18). In addition, alpha satellite from a CENP-B box rich region of the chromosome 21 centromere (21-I) forms *de novo* centromeres in an SMC system, while alpha satellite from a neighboring CENP-B box depleted region (21-II) is inefficient (19). Further, the *de novo* centromere nucleation ability of the 21-I-derived alpha satellite array can be disrupted by mutation of its constituent CENP-B boxes (15), an outcome that parallels our observations on mutation of CENP-B boxes in D17Z1-derived alpha satellite. Finally, it has also been established that CENP-B boxes outside the context of alpha satellite DNA are not competent to nucleate *de novo* centromere assembly (15), establishing that sequence features other than CENP-B boxes are also required for centromere function. Taken together, our data and the earlier observations unambiguously establish the presence of CENP-B and its cognate binding element as a requirement for efficient *de novo* centromere formation in SMC or artificial chromosome assays.

Notwithstanding the clear role of the CENP-B box in assembly of SMCs, the role of CENP-B in its endogenous chromosomal context remains open to debate. At least three CENP-B-like proteins have been identified in fission yeast, and double mutants exhibit severe chromosome segregation defects (33). Such functional redundancy may explain the lack of a major phenotype in mouse knockouts of CENP-B (29, 30, 31) and why CENP-B appears dispensable for

function of the Y chromosome in both mice and men, as well as for function of neocentromeres and certain dicentric chromosomes (34, 35). In addition, it remains to be established whether the position of CENP-B boxes within an array of monomers or even within a single monomer is also of importance, as might be expected if CENP-B participates in nucleosome positioning (36, 37).

In addition to the effect of manipulating CENP-B boxes demonstrated here and by Ohzeki et al. (15), it is apparent that other sequences within alpha satellite may influence the efficiency of SMC formation, as even arrays with a similar number of CENP-B boxes can differ quite substantially in their ability to seed SMCs (Rudd et al., in press; 25). This possibility may now be investigated systematically using synthetic alpha satellite arrays where the distribution of CENP-B boxes and/or other sequences in each monomer has been manipulated, using the approach outlined here. Determination of the ideal density and distribution of such sequences in alpha satellite will maximize the efficiency with which SMC vectors carrying therapeutic genes might eventually be assembled in human cells (14, 7, 12).

TABLE ONE

Effect of CENP-B box density on efficiency of SMC formation

Construct	**CENP-B box density**	**Experiments (no.)**	**Clones screened (no.)**	**Clones with SMC (no.)**	**SMC formation frequency**
Natural D17Z1	5/16	6	38	4	10.5 %
All CENP-B+	16/16	15	45	10	22 %
CENP-B null	0/16	10	40	1	2.5 %

FIGURE LEGENDS

Figure 1. (A) Outline of iterative scheme for synthesis of mutant versions of chromosome 17 alpha satellite arrays. Each of the 16 individual monomers comprising a single higher-order repeat (HOR) was synthesized as 2-3 oligonucleotide pairs (60-100 bp each), which were directly ligated together and gel purified. Adjacent repeat units were subsequently ligated to form dimers as shown and PCR-modified to introduce SapI recognition sites at both ends as appropriate. Digestion with SapI allows seamless ligation of adjacent dimers to create tetramers without introduction of extraneous non-alpha satellite sequences. Two additional rounds of serial ligation resulted in formation of a complete synthetic higher-order repeat unit, which was subcloned into the BAC vector pBeloBAC (Shizuya et al., 1992), creating pBAC17α1(all CENP-B+/all CENP-B-).

(B) Outline of scheme for directional multimerization of engineered higher-order repeats. A synthetic alpha satellite array consisting of 32 tandemly multimerized copies of the higher-order repeat was created as follows: pBAC17α1 was digested with BamHI and SpeI and the alpha satellite containing fragment (fragment 'A') isolated and gel purified. The same construct was separately digested with BglII and SpeI, and the larger fragment (fragment 'B') isolated and gel purified. Ligation of fragment 'A' to fragment 'B' is directional, resulting in head-to-tail multimerization of adjacent higher-order repeats. The resulting pBAC17α2 construct was then isolated following transformation of the ligation reaction into *E.coli.* This process was repeated iteratively to create the final pBAC17α32 arrays.

(C) Pulsed Field Gel Electrophoresis (PFGE) analysis of intermediates in the construction of 17α32 HOR/BeloBAC constructs. Each intermediate was digested

with NotI, which excises the entire subcloned alpha satellite array from the pBeloBAC vector backbone. Lanes are labeled according to higher-order repeat copy number. The insert in lane 4 is 2.7 kb and therefore too small to be resolved by PFGE.

Figure 2. Mobility shift analysis of synthetic CENP-B box enriched and CENP-B box null monomers. Ligated tetramers of CENP-B box-enriched and CENP-B box-null monomers were electrophoresed through an agarose gel following incubation with purified recombinant CENP-B protein. Lanes 1, 2, and 3 represent enriched tetramers, while lanes 4, 5, and 6 contain null species. Tetramer DNAs (100ng) were pre-incubated with varying quantities of CENP-B protein for 25 minutes at room temperature and subsequently loaded into a 2% agarose gel. Lanes 2 and 5 (20μg protein) as well as lanes 3 and 6 (40μg protein) contain protein/DNA mixtures. Comparison of lanes 2 and 3 to lanes 5 and 6 reveals an marked difference in mobility shift in the CENP-B box-enriched subunits, while only a modest shift is seen with CENP-B box-null DNA. This slight mobility shift is likely due to salt effects as similar results are observed with a buffer-only control (data not shown).

Figure 3. Cytogenetic detection of SMCs from synthetic chromosome 17-derived alpha satellite arrays. Arrows designate SMCs. Immunostaining with an anti-CENP-C antibody (green) identifies functional centromeres. FISH with the synthetic alpha satellite as probe (red) hybridizes with the synthetic microchromosome as well as to the centromeres of the endogenous chromosome 17s. DAPI stained DNA is shown in blue.

(A) HT1080 clone generated by transfection with pBAC17α32(All CENP-B+), showing the presence of two SMCs.

(B) HT1080 clone generated by transfection with pBAC17α32(natural). A single SMC is visible.

(C) HT1080 clone generated by transfection with pBAC17α32(CENP-B null). Two putative SMCs are present in this clone, but none were detected in all other clones obtained with the CENP-B null construct.

ACKNOWLEDGMENTS

We are grateful to Dr. Huntington Willard (Institute for Genome Sciences and Policy, Duke University) for access to data prior to publication and for providing critical review of this manuscript. We thank Linda Woods for performing the cytogenetic analysis.

REFERENCES

1. Sullivan BA, Blower MD, Karpen GH. (2001) Determining centromere identity: cyclical stories and forking paths. *Nat Rev Genet.*, 2(8):584-96

2. Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC. Human centromeric DNAs. (1997) *Hum Genet.*, 100; 291-304

3. Choo KH, Vissel B, Nagy A, Earle E, Kalitsis P. (1991) A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res.*, 19(6):1179-82.

4. Willard HF. (1991) Evolution of alpha satellite. *Curr Opin Genet Dev.*, 1(4):509-14

5. Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC. (1997) The centromere: hub of chromosomal activities. *Science*, 270; 1591-1594

6. Amor DJ, Choo KH. (2002) Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet.*, 71(4):695-714

7. Saffery R and Choo KH. (2002) Strategies for engineering human chromosomes with therapeutic potential. *J. Gene Med.*, 4; 5-13

8. Grimes BR, Rhoades AA, Willard HF. (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol Ther.*, 5(6):798-805

9. Willard HF. (2001) Neocentromeres and human artificial chromosomes: an unnatural act. *Proc Natl Acad Sci U S A.*, 98(10):5374-6

10. Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF. (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet.*, 15(4):345-55

11. Waye JS, Willard HF. (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal

crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol.*, 6(9):3156-65

12. Willard HF. (2000) Genomics and gene therapy. Artificial chromosomes coming to life. *Science*, 17;290(5495):1308-9

13. Mejia JE, Willmott A, Levy E, Earnshaw WC, Larin Z. (2001) Functional complementation of a genetic deficiency with human artificial chromosomes. *Am J Hum Genet.*, 69(2):315-26.

14. Grimes BR, Schindelhauer D, McGill NI, Ross A, Ebersole TA, Cooke HJ (2001) Stable gene expression from a mammalian artificial chromosome. *EMBO Rep.* 2(10):910-4.

15. Ohzeki J, Nakano M, Okada T, Masumoto H. (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol.*, 159(5):765-75

16. Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF. (2001) Genomic and genetic definition of a functional human centromere. *Science.* 294(5540):109-15.

17. Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat Biotechnol.*, 16(5):431-9.

18. Mejia JE, Alazami A, Willmott A, Marschall P, Levy E, Earnshaw WC, Larin Z. (2002) Efficiency of de novo centromere formation in human artificial chromosomes. *Genomics*, 79(3):297-304

19. Masumoto H, Ikeno M, Nakano M, Okazaki T, Grimes B, Cooke H, Suzuki N. (1998) Assay of centromere function using a human artificial chromosome. *Chromosoma*, 107(6-7):406-16

20. Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol.*, 116(3):585-96

21. Cooke CA, Bernat RL, Earnshaw WC. (1990) CENP-B: a major human centromere protein located beneath the kinetochore. *J Cell Biol* 110(5):1475-88

22. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol.*, 109(5):1963-73

23. Earnshaw WC, Sullivan KF, Machlin PS, Cooke CA, Kaiser DA, Pollard TD, Rothfield NF, Cleveland DW. (1987) Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *J Cell Biol.*, 104(4):817-29

24. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A.*, 89(18):8794-7

25. Kouprina N, Ebersole T, Koriabine M, Pak E, Rogozin IB, Katoh M, Oshimura M, Ogi K, Peredelchuk M, Solomon G, Brown W, Barrett JC, Larionov V. (2003) Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. *Nucleic Acids Res.*, 31(3):922-34.

26. Grimes BR, Warburton PE, Farr CJ. (2002) Chromosome engineering: prospects for gene therapy. *Gene Ther.*, 9(11):713-8

27. Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y. (2001) Alpha-satellite DNA of primates: old and new families. *Chromosoma,* 110(4):253-66.

28. Kipling D, Warburton PE. (1997) Centromeres, CENP-B and Tigger too. *Trends Genet.*, 13(4):141-5.

29. Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC. (1996) Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. *Mol Cell Biol.* 16(9):5156-68.

30. Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, Mancini M, Ouspenski I, Brinkley BR, May GS. (1998) The cenpB gene is not essential in mice. *Chromosoma.*, 107(8):570-6

31. Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL, Vogel KG, Moyzis RK, Pedersen R. (1998) Centromeric protein B null mice are viable with no apparent abnormalities. *Dev Biol.*, 201(2):135-43

32. Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, Hill J, Wreford NG, de Kretser DM, Cancilla MR, Howman E, Hii L, Cutts SM, Irvine DV, Choo KH. (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol.*, 141(2):309-19

33. Irelan JT, Gutkin GI, Clarke L. (2001) Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics*, 157(3):1191-203.

34. Choo KH. (1997) Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am J Hum Genet.*, 61(6):1225-33

35. Earnshaw WC, Ratrie H 3rd, Stetten G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma*, 98(1):1-12.

36. Yoda K, Ando S, Okuda A, Kikuchi A, Okazaki T. (1998) In vitro assembly of the CENP-B/alpha-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. *Genes Cells*, 3(8):533-48

37. Warburton PE, Waye JS, Willard HF. (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol Cell Biol.* 13(10):6520-9.

Incorporation By Reference:

All references cited throughout the specification and listed below are incorporated herein by reference.

- US Patent No. 5,695,967, Van Bokkelen et al., relating to a method for stably cloning large repeating units of DNA.
- US Patent No. 5,869,294, Harrington et al., relating to a method for stably cloning large repeating units of DNA.
- US Patent No. 6,348,353, Harrington et al., relating to constructing artificial (synthetic) mammalian chromosomes.

Definitions:

"Higher Order Repeat DNA" refers to a repeating unit that is itself composed of smaller (monomeric) repeating units. The basic organizational unit of alpha satellite arrays is the approximately 171 base pair alphoid monomer. Monomers are organized into chromosome-specific higher order repeating units, which are also tandemly repetitive. The number of constituent monomers in a given higher order repeat varies, from as little as two (for example, in human chromosome 1) to greater than 30 (human chromosome Y). Constituent monomers exhibit varying degrees of homology to one another, from approximately 60% to virtual sequence identity. However, higher order repeats retain a high degree of homology throughout most of a given alphoid array.

"Synthetic" refers to a molecule that does not naturally occur in nature and/or has been constructed *de novo* by man. For example, a minichromosome constructed by the recombination and/or breakage of a natural chromosome is not a synthetic chromosome.

"Seamless" restriction enzyme refers to any restriction enzyme that would allow ligation of two DNA fragments of a higher repeat order DNA (such as the pairs of adjacent dimers shown in Figure 1A) to form a larger fragment (such as the tetramers shown in Figure 1A) without introduction of extraneous non-alpha satellite sequences. Examples of "Seamless" enzymes include the class of restriction enzymes known as TypeIIS. TypeIIS enzymes like FokI and AlwI cleave outside of their recognition sequence to one side. These enzymes are intermediate size, typically 400-650 amino acids in length, and they recognize sequences that are continuous and asymmetric. They comprise two distinct domains, one for DNA binding, the other for DNA cleavage. They are thought to bind to DNA as monomers for the most part, but to cleave DNA cooperatively, through dimerization of the cleavage domains of adjacent enzyme molecules. For this reason, some Type IIS enzymes are much more active on DNA molecules that contain multiple recognition sites.

"Isoschizomer" refers to a restriction enzyme that recognizes the same nucleotide sequence as another restriction enzyme and cleaves that same sequence. Therefore, a **"Non-isoschizomeric site"** refers to a restriction enzyme site that can be cut by one of two restriction enzymes, but not by both.

"Stably transformed" refers to the fact that a cloned DNA array containing the repeating units is capable of being propagated in the desired host cell for at least 50 generations of growth with a recombination frequency of less than 0.6% per generation (for 174 kb arrays) and a recombination frequency of less than 0.2% (for 130 kb arrays).

"Directionally" as in **"directionally ligating"** refers to the order of the fragments that are ligated together in a sequential order, following the sequence of the DNA unit that is being constructed. For example, in constructing a fragment with the following sequence "ATTTTTTAGCGCCCGGTTTATTTACCCCCCCC," the smaller fragments that are first constructed span the full length of the larger fragment. For example, 4 smaller fragments may be constructed with the following sequences: Fragment 1 = ATTTTTTA;

Fragment 2 = GCGCCCGG; Fragment 3 = TTTATTTA; and Fragment 4 = CCCCCCCC. By "directionally ligating" the smaller fragments, therefore, it is meant that small fragment 1 is ligated to small fragment 2 and the small fragment 3 is ligated to small fragment 4, all in the same sequential orientation 5' to 3' or 3'to 5', to maintain the sequence of the larger fragment that is to be constructed. It would NOT be "directionally ligating" if fragment 1 were to be ligated to fragment 3 or 4 and/or if the 5' to 3' direction of the sequence of any one small fragment was disrupted (as in ligating the small fragment 1 in its 5'-3' direction to the small fragment 2 in its 3'-5' direction, resulting in a larger fragment with the sequence ATTTTTTA + GGCCCGCG, instead of the directional ligation sequence of ATTTTTTA + GCGCCCGG).

Clarifications:

"non-naturally occurring distribution of CENP-B boxes," as appears for example in the claims, refers to the fact that not only the number of CENP-B boxes on a given chromosome may vary but also the distribution of the CENP-B boxes vary. In the present invention, both the distribution of the CENP-B boxes as well as the number of CENP-B boxes may be altered to form a desired DNA construct. For example, a construct may contain a CENP-B box in every HOR or one in every other HOR, or none in the first 5 HOR, and so on and so forth. Such constructs are useful *per se* (as for example, increasing efficiency of SMC formation) or useful in a variety of ways in the elucidation of the role of various permutations in centromere formation and function.

Figure 1B:

This is further clarification of Figure 1B:

The construct pBeloBAC17alpha X HOR CENP-B box saturated/null is the starting vector. X is the number of copies of the HOR in a given iteration. X may equal 1, 2, 4, 8, 16, 32, etc.
Taking the embodiment where X = 1, as shown in Figure 1B, digestion of the starting construct with BamH1 and Spe1 creates an insert fragment, referred to as "A," consisting of the HOR plus a small amount of vector sequence. Digestion of the starting construct with Bgl2 and Spe1 creates the corresponding vector fragment or "B," consisting of the starting vector minus the small amount of sequence between the Bgl2 and Spe1 sites. A is now cloned into B to give the pBeloBAC17alpha2HOR, shown on the right, in Figure 1B. Reiteration of this process builds up the array to pBeloBAC17alpha32HOR and so forth.

What we Claim is:

1. A synthetic higher order repeat DNA with non-naturally occurring distribution of CENP-B boxes.

2. A synthetic higher order repeat DNA enriched in CENP-B box sequences.

3. The synthetic higher order repeat DNA of claim2, wherein the number of CENP-B Boxes of the synthetic higher order repeat DNA is greater than the number of CENP-B boxes of its counterpart naturally occurring higher order repeat DNA.

4. A synthetic alpha satellite DNA comprising a higher order repeat as claimed in claims 1-3.

5. A synthetic microchromosome vector containing the alpha satellite array of claim 4.

6. A synthetic microchromosome formed by introduction of the synthetic microchromosome vector of claim 5 into an appropriate cell.

7. The synthetic microchromosome vector of claim 5, wherein said vector when introduced in an appropriate cell forms a synthetic microchromosome at an efficiency rate higher than a microchromosome vector containing a higher order repeat DNA with an unaltered CENP-B box frequency and distribution.

8. A method of increasing efficiency of formation of a synthetic michrochromosome comprising constructing a synthetic microchromosome vector containing one or more higher order repeat DNA as claimed in claims 1-3; and introducing said synthetic microchromosome vector into an appropriate cell, thereby forming a synthetic microchromosome.

9. A method of making a synthetic repetitive DNA array comprising:
   (a) constructing a synthetic monomer of defined DNA sequence;
   (b) directionally assembling said synthetic monomers to form the desired synthetic repetitive DNA array.

10. The method of claim 9, wherein said synthetic repetitive DNA array is a higher order repeat DNA.

11. A synthetic repetitive DNA array made by the process of claim 9.

12. A higher order repeat DNA made by the method of claim 10.

13. A method of synthesizing a desired higher order repeat DNA comprising:

(a) synthesizing each monomer unit of said desired higher order repeat DNA as one or more oligonucleotide(s);
(b) directionally ligating pairs of adjacent monomer units to form repeating monomeric units to form the desired higher order repeat DNA.

14. A higher order repeat DNA made by the method of claim 13.

15. A method of synthesizing a desired higher order repeat DNA comprising:
(a) synthesizing each monomer unit of said desired higher order repeat DNA as one or more oligonucleotide pairs which are directly ligated together to form a synthetic monomer unit;
(b) directionally ligating pairs of adjacent monomer units to form dimers;
(c) modifying said dimers to introduce "seamless" restriction enzyme recognition sites at both ends of each dimer, thereby forming modified dimers;
(d) digesting said modified dimers with a "seamless" restriction enzyme which cuts said modified dimmers at said "seamless" restriction enzyme recognition site, thereby forming dimers with "seamless" overhangs;
(e) directionally ligating pairs of adjacent dimers with "seamless" overhangs, thereby forming tetramers;
(f) repeating modification, digestion, and directional ligation as set forth in (c)-(e), above, until all the monomer units of the desired higher order repeat DNA are ligated together in two separate groups, forming two multimers;
(g) ligating said two multimers to form desired higher order repeat DNA.

16. A higher order repeat DNA made by the process of claim 15.

17. A method of making a synthetic alpha satellite array comprising:
(a) modifying a first higher order repeat DNA of any one of claims 12, 14, or 16 such that the opposing ends of the higher order repeat DNA contain complementary, but non-isoschizomeric restriction enzyme sites, thereby forming a modified higher order repeat DNA;
(b) ligating said modified higher repeat DNA into a vector that contains the same isoschizomeric restriction enzyme sites as the ends of said modified higher order repeat DNA;
(c) lineraizing said vector at one of said isoschizomeric restriction enzyme sites;
(d) ligating into said vector, in tandem with said first modified higher repeat DNA, a second higher order repeat DNA of either

claims 1 or 8 modified in the same way as the first higher order repeat DNA in (a); so as to form a directional repeating array;

(e) transforming said directional repeating array into a bacterial host cell;

(f) selecting stable clones containing said directional repeating array; and

(g) repeating (c)-(f) until a desired alpha satellite array size is reached.

18. A synthetic microchromosome vector comprising one or more higher order repeat DNA as claimed in claims 12, 14, and 16.

19. The synthetic microchromosome vector of claim 18, wherein said vector when introduced in an appropriate cell forms a synthetic microchromosome at an efficiency rate higher than a microchromosome vector containing a higher order repeat DNA with an unaltered CENP-B box frequency and distribution.

20. A synthetic microchromosome formed by introduction of the synthetic microchromosome vector of claims 18 or 19 into an appropriate cell.

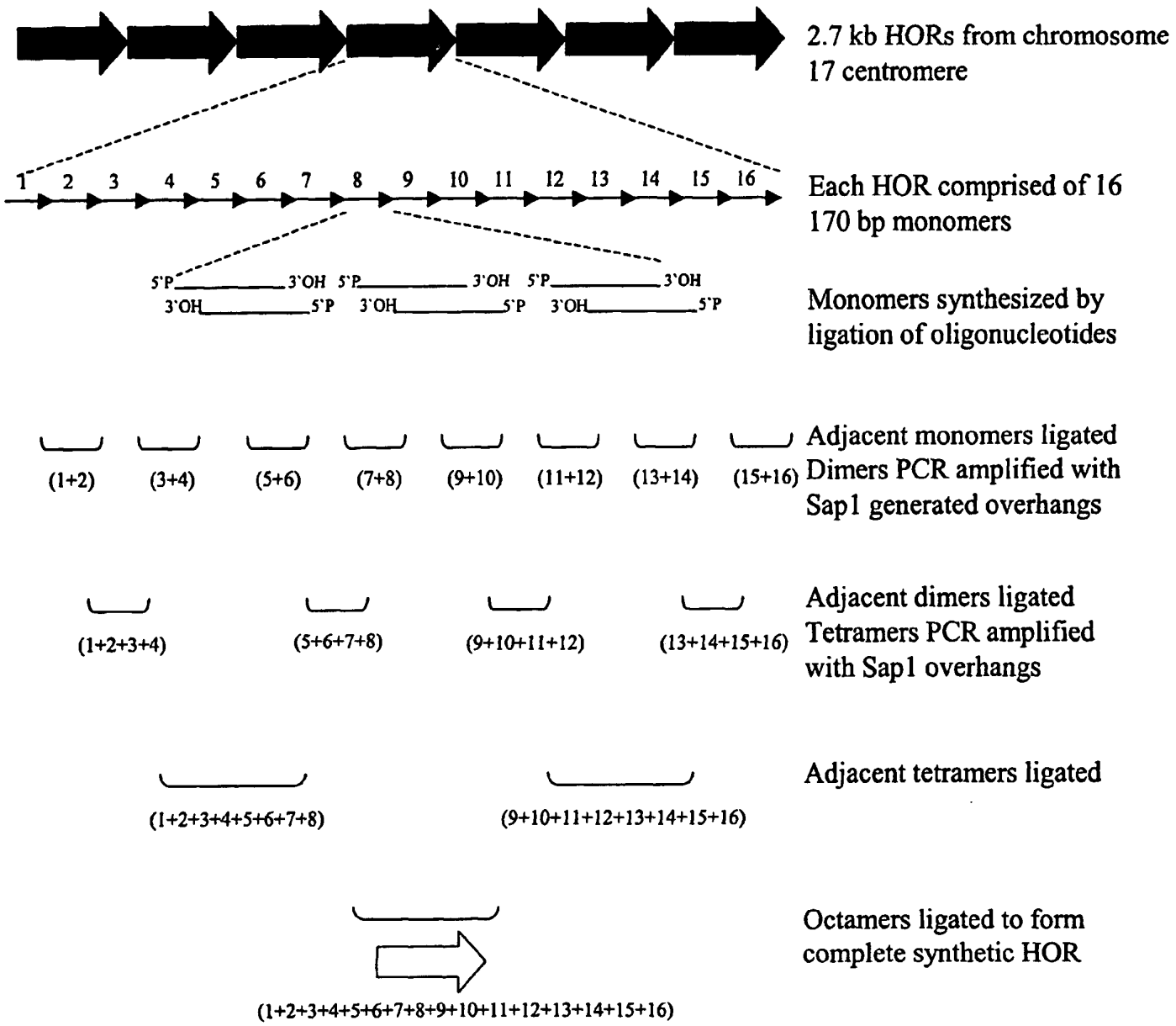

2.7 kb HORs from chromosome 17 centromere
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Each HOR comprised of 16 170 bp monomers
5'P 3'OH 5'P 3'OH 5'P 3'OH
3'OH 5'P 3'OH 5'P 3'OH 5'P
Monomers synthesized by ligation of oligonucleotides
(1+2) (3+4) (5+6) (7+8) (9+10) (11+12) (13+14) (15+16)
Adjacent monomers ligated Dimers PCR amplified with Sap1 generated overhangs
(1+2+3+4) (5+6+7+8) (9+10+11+12) (13+14+15+16)
Adjacent dimers ligated Tetramers PCR amplified with Sap1 overhangs
(1+2+3+4+5+6+7+8) (9+10+11+12+13+14+15+16)
Adjacent tetramers ligated
(1+2+3+4+5+6+7+8+9+10+11+12+13+14+15+16)
Octamers ligated to form complete synthetic HOR

Figure 1A

1. Digest "insert" with BamHI and SpeI: fragment 'A'
2. Digest "vector" with Bgl II and SpeI: fragment 'B'
3. Ligate, transform *E.coli* and identify pBeloBAC 17α32 2(HOR)
4. Repeat iteratively

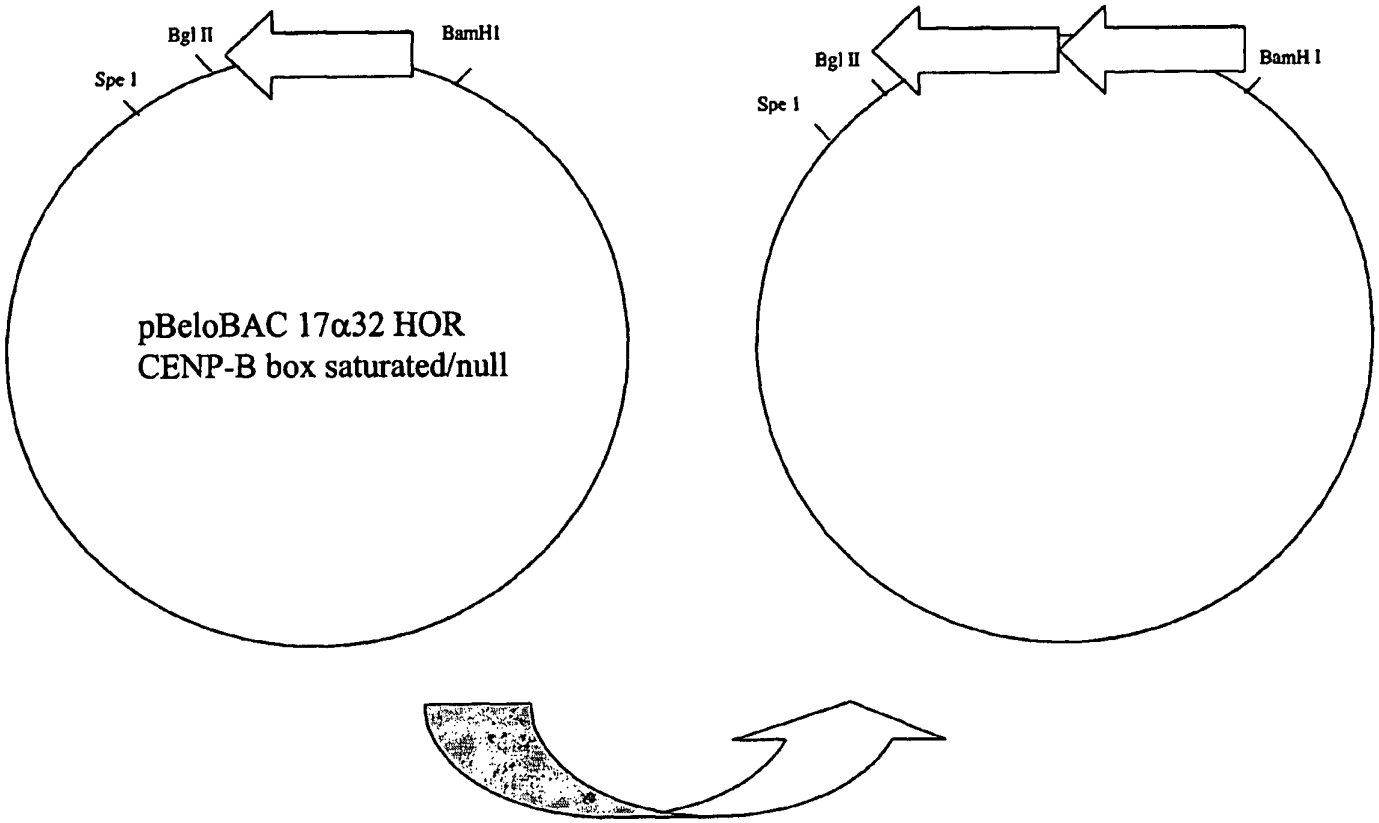


Figure 1B

L

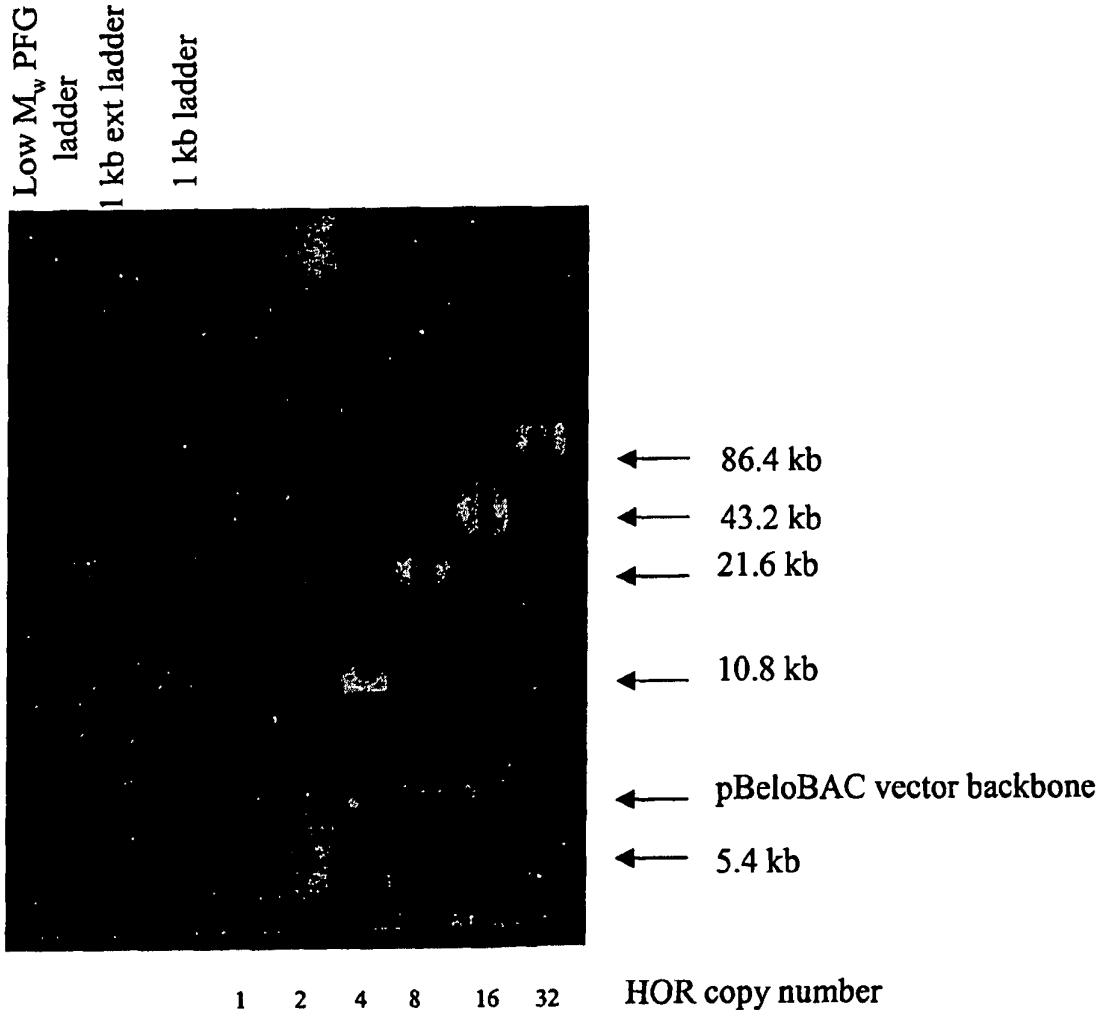


Figure 1C

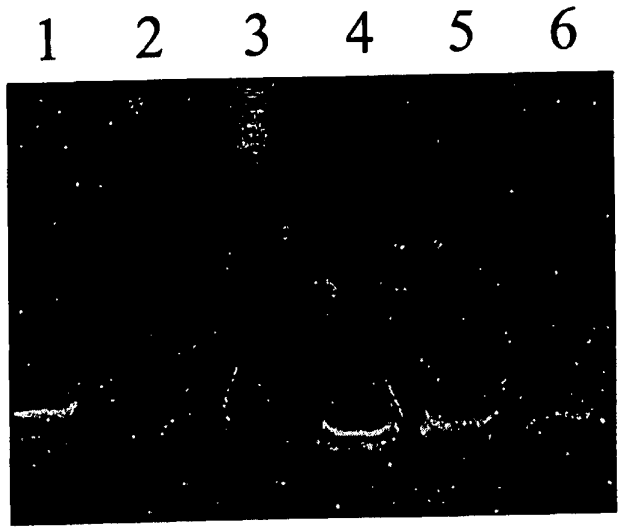
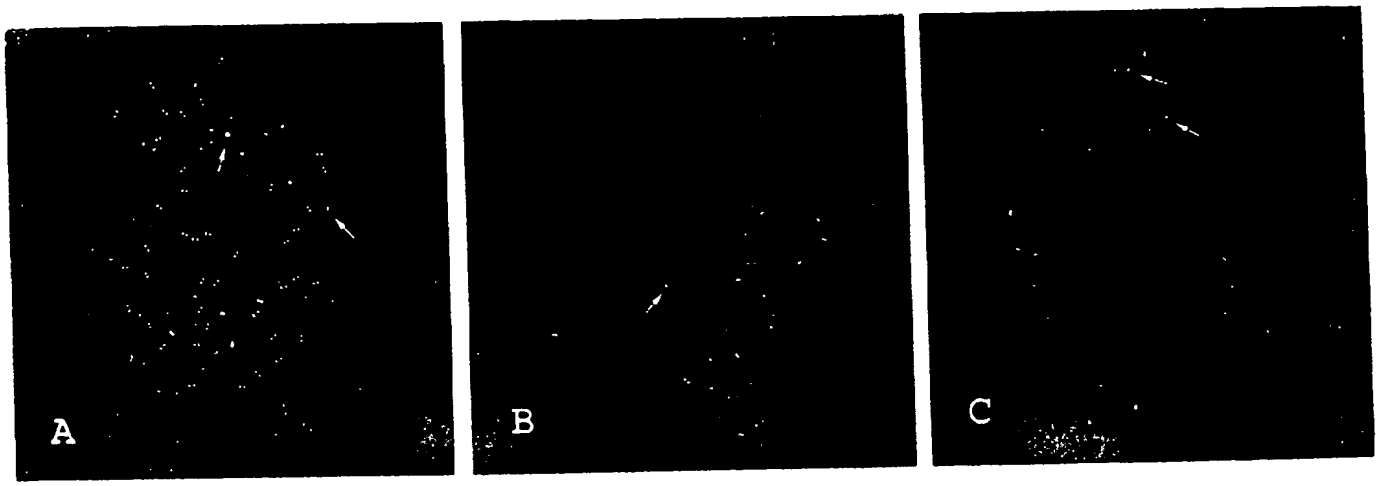


Figure 2



Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/022314

International filing date: 09 August 2004 (09.08.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/492,965
Filing date: 07 August 2003 (07.08.2003)

Date of receipt at the International Bureau: 12 November 2004 (12.11.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse